PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference GB020064	**FOR FURTHER ACTION**	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/GB 03/04589	International filing date *(day/month/year)* 22.10.2003	Priority date *(day/month/year)* 10.04.2003
International Patent Classification (IPC) or both national classification and IPC G06F11/14, G06F17/30		
Applicant INTERNATIONAL BUSINESS MACHINES CORPORATION et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 6 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 10 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the opinion
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
10.11.2003	16.02.2005
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized Officer Lanchès, P Telephone No. +49 89 2399-7440

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB 03/04589

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:

Description, Pages

1, 2, 5-23	as originally filed
3, 4, 4a	received on 02.12.2004 with letter of 01.12.2004

Claims, Numbers

1-16	received on 02.12.2004 with letter of 01.12.2004

Drawings, Sheets

1/5-5/5	as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

   These elements were available or furnished to this Authority in the following language: , which is:

   ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).

   ☐ the language of publication of the international application (under Rule 48.3(b)).

   ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

   ☐ contained in the international application in written form.

   ☐ filed together with the international application in computer readable form.

   ☐ furnished subsequently to this Authority in written form.

   ☐ furnished subsequently to this Authority in computer readable form.

   ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

   ☐ the description, pages:

   ☐ the claims, Nos.:

   ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)).

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-16
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-16
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-16
	No:	Claims	

2. Citations and explanations

   see separate sheet

V Reasoned statement under Article 35(2)

1 Prior art

Reference is made to the following document:

D1: WO 02/073409 A (ERICSSON TELEFON AB L M ; RONSTROEM MIKAEL (SE)) 19 September 2002 (2002-09-19)

2 Objections under Article 6 PCT

The present application does not meet the requirements of Article 6 PCT, because an essential feature is missing in the independent claims.

The present invention only makes sense for systems in which it is not necessary to process data items in the same order as they were added to the repository (see the description page 4, lines 39-41, page 8, lines 13-16 and page 12, lines 28-39). This is an important restriction which constitutes an essential feature of the invention, and which should, therefore, have been included in all the independent claims (PCT Guidelines, section 5.29). In particular, no inventive step can be acknowledged without this feature, see section V-3.4.

3 Article 33 PCT

3.1 The following reasoning only applies provided that the above mentioned essential feature is incorporated in the independent claims.

Document D1 is considered to represent the closest prior art with respect to the subject-matter of independent claim 1. It discloses (the references in parentheses applying to D1):

A method for recovering a data repository from a failure affecting a primary copy of the data repository, including the steps of:

(page 5, lines 10-18)

maintaining a secondary copy of data sufficient to recover the primary copy of the data repository and data items held thereon;

in response to a failure affecting the primary copy of the data repository, recreating a primary copy of the data repository from the secondary copy; and using a restore process to restore data items to the primary copy from the secondary copy

(page 5, lines 22-28)

within a recovery unit of work, wherein data items restored to the primary copy of the data repository within the recovery unit of work are made inaccessible to processes other than the restore process until commit of the recovery unit of work;

(page 16, lines 1-4)

prior to commit of the recovery unit of work, configuring the primary copy of the data repository to enable addition of data items to the data repository independent of said restore step;

(page 12, lines 1-6)

in response to successful completion of the restore step, committing the recovery unit of work including releasing said inaccessibility of the restored data.

(page 16, lines 4-5)

3.2 The difference between the claimed invention and the closest prior art is:

configuring the primary copy of the data repository to enable processes other than the restore process to retrieve said independently added data items.

Hence, the subject-matter of independent claim 1 is novel.

3.3 The objective technical problem to be solved by the claimed invention. therefore, is:

optimise the performance during restore processing after a failure of the primary repository.

3.4 The technical feature identified under V-3.2 provides a non-obvious solution to said problem in that it allows retrieve operations to be performed both on the primary repository being restored and on the secondary repository, from which the restore is performed. Such processing is possible and useful in systems in which it is not necessary to process data items in the same order as they were added to the repository. Document D1 explicitly excludes read transactions on the restored repository during the restore operation. However, this restriction only results from the necessary ordered processing of the log in D1, which is necessary for applications in general. This no longer is a requirement in the present invention. Without this requirement, however, a modification of the system according to D1 falling under the terms of present independent claim 1 would be an obvious alternative, since reading the restored items available on the repository under restoration, clearly, is already technically possible there. Hence, the essential feature mentioned in section V-2 is a distinction from the disclosure of D1 which is necessary for the acknowledgement of an inventive step.
Under these circumstances, the subject-matter of independent claim 1 involves an inventive step.

3.5 The industrial applicability of the present invention is obvious.

3.6 The reasoning of sections V-3.1 to V-3.5 applies similarly to corresponding independent system claim 13, corresponding independent computer program product claim 15 and corresponding independent computer program claim 16.

3.7 Consequently, the subject-matter of the remaining dependent claims also fulfills the requirements of Article 33 PCT.

4 Objections under Article 6 PCT

The present application does not meet the requirements of Article 6 PCT, because dependent claim 14 is not concise and because said claim is not clear:
Said claim 14 depends on present claim 13 and simultaneously appears to redundantly redefine features already introduced in claims 13. Instead. it should simply have specified the message specific aspects of the repository. and the corresponding processing.

5 Certain defects in the international application

5.1 The Independent claims are not in the two-part form in accordance with Rule 6.3(b) PCT, which in the present case would have been appropriate, with those features known in combination from the prior art document D1 being placed in the preamble (Rule 6.3(b)(I) PCT) and with the remaining features being included in the characterising part (Rule 6.3(b)(ii) PCT).

5.2 The features of the claims are not provided with reference signs placed in parentheses (Rule 6.2(b) PCT). This applies to both the preamble and the characterising portion (see the PCT Guidelines, III-4.11).

transactions that have already been applied to the target database during recovery.

An alternative approach is described in US Patent No. 6,353,834 issued on 5 March 2002 to Wong et al, in which a message queueing system stores messages and state information about the messages, clustered together in a single file on a single disk. This system is intended to achieve efficient writing of data by avoiding writing updates to three different disks (a data disk, an index structure disk and a log disk). A Queue Entry Map Table is used to enter control information, message blocks and log records. US 6,353,834 refers to the use of existing RAID technology and duplicate writing of data, without which the described system provides no protection against storage failures which result in loss of the data held on the single disk.

International Patent Application Publication Number WO 02/073409 discloses a method for recovery of database nodes without stopping write transactions. A failed node is restored using an old version of a database fragment in the failed node together with an up-to-date version of the fragment in another node, by copying the parts of the fragment which have changed since creation of the old version. A delete log is used to enable the recovery processing to take account of deletions since the creation of the old version. Write transactions occurring after the start of recovery processing are performed on the recovering node during the recovery processing.

Summary

Aspects of the present invention provide methods, data processing systems, recovery components and computer programs for recovering from failures affecting data repositories, wherein at least a part of the recovery processing is performed while the data repositories are able to receive new data and to allow retrieval of such new data. The failure may be a hardware failure or malfunction, or a software malfunction, which results in loss or corruption of data in a data repository on a primary storage medium.

Although new data items (i.e. those received after the failure) may be received into the repository and retrieved therefrom during recovery processing, updates to the data repository which were performed before the failure and which are then restored to the repository by the recovery processing are made inaccessible until completion of the recovery

processing. The recovery processing can achieve fast availability of the data repository while also ensuring that the recovered repository is consistent with the state of the repository at the time of the failure.

In a first aspect, the present invention provides a method for recovering a data repository from a failure affecting a primary copy of the data repository, including the steps of: maintaining a secondary copy of data sufficient to recreate the primary copy of the data repository and data items held thereon; in response to a failure affecting the primary copy of the data repository, recreating a primary copy of the data repository from the secondary data copy, and using a restore process to restore data items to the primary copy from the secondary copy within a recovery unit of work; wherein data items restored to the primary copy of the data repository within the recovery unit of work are made inaccessible to processes other than the restore process until commit of the recovery unit of work; prior to commit of the recovery unit of work, configuring the primary copy of the data repository to enable addition of data items to the data repository independent of said restoring step and to enable processes other than the restore process to retrieve said independently added data items; and in response to successful completion of the restoring step, committing the recovery unit of work including releasing said inaccessibility of the restored data.

According to a preferred embodiment of the invention, updates to a message repository during normal forward processing of a messaging system include message send operations which add messages to the repository, and message retrieve operations which delete the messages. The 'message repository' in this context may be a message queue, a database table, or any other data structure which holds messages or message queues. Following a failure which affects the message repository, the message repository is recreated in an empty state and then send and retrieve operations are reapplied to the repository, preferably by referring to a backup copy of the repository and log records. The message repository is recreated as a preliminary recovery step and messaging functions are able to transfer new messages to and from the message repository prior to completion of recovery. Updates to the message repository which involve reapplying operations from backup storage and log records are handled as uncommitted operations of a Recovery Unit of Work and only committed (i.e. a consistency check is performed and the updates are made final and accessible to other programs) on completion of the Recovery Unit of Work. The Recovery Unit of Work includes the set of operations required

(following recreation of the message repository) to restore the contents of the message repository to a state consistent with the state of the repository at the time of the failure. The message repository is available for receipt of new messages as soon as it is recreated, whereas any message which is restored to a queue within the Recovery Unit of Work cannot be retrieved from the repository by a target application program until completion of the Recovery Unit of Work.

The invention is useful for applications in which it is not essential to process data items in the same order as they were added to the data repository. In a first example application, each data item or message is a request for performance of a particular task. If the order of

CLAIMS

1. A method for recovering a data repository from a failure affecting a primary copy of the data repository, including the steps of:

maintaining a secondary copy of data sufficient to recover the primary copy of the data repository and data items held thereon;

in response to a failure affecting the primary copy of the data repository, recreating a primary copy of the data repository from the secondary copy; and

using a restore process to restore data items to the primary copy from the secondary copy within a recovery unit of work, wherein data items restored to the primary copy of the data repository within the recovery unit of work are made inaccessible to processes other than the restore process until commit of the recovery unit of work;

prior to commit of the recovery unit of work, configuring the primary copy of the data repository to enable addition of data items to the data repository independent of said restore step and to enable processes other than the restore process to retrieve said independently added data items; and

in response to successful completion of the restore step, committing the recovery unit of work including releasing said inaccessibility of the restored data.

2. A method according to claim 1, wherein maintaining the secondary data copy comprises storing a backup copy of the data repository and storing log records describing updates to the primary copy performed since the backup copy was stored; wherein recreating the primary copy of the data repository includes the step of copying data repository definitions from the backup copy and applying the definitions to recreate the primary copy; and wherein restoring data items to the primary copy comprises copying data items from the backup copy and replaying the log records to identify and reapply updates to the primary copy.

3. A method according to claim 1, wherein maintaining the secondary data copy includes storing log records that describe updates to the primary copy, and wherein the step of restoring the primary copy of the repository includes the steps of:

replaying the log records of operations performed on data items within the primary copy of the data repository,

caching log records relating to operations performed under syncpoint control within an original unit of work,

determining from the cached log records the state of the original units of work at the time of the failure, and

determining which of said syncpoint-controlled operations to perform within the recovery unit of work based on the determined state of the original units of work.

4. A method according to claim 3, including performing operations within the recovery unit of work in accordance with the following procedure:

if the original unit of work was committed before the failure, performing the relevant operations of the committed unit of work;

if the original unit of work was in-doubt when the failure occurred, performing the relevant operations of the in-doubt unit of work but marking the operations in-doubt; and

if the original unit of work is neither committed nor in-doubt, discarding the cached operations.

5. A method according to claim 3, including discarding from the recovery unit of work any pairs of addition and deletion operations that comprise an addition of a data item to the primary copy of the data repository and a deletion of the same data item from the primary copy of the data repository, on condition that said addition and deletion operations were performed and committed before the failure.

6. A method according to any one of the preceding claims, wherein the data repository is a message repository and the step of restoring data to the primary copy of the data repository comprises performing message add, update and delete operations on the message repository.

7. A method according to claim 6, for performance within a messaging communication system, wherein maintaining the secondary data copy includes storing log records to describe updates to the primary copy, and wherein the step of restoring data to the primary copy of the repository includes

the steps of caching log records relating to message add, update and delete operations performed under syncpoint control within an original unit of work, determining from the log records the state of the original unit of work at the time of the failure, and determining the operations to perform within the recovery unit of work based on the determined state of the original unit of work as follows:

if the original unit of work is committed, performing the relevant message add, update and delete operations; and

if the original unit of work is in-doubt, performing the relevant message add, update and delete operations but marking the operations in-doubt; and

if the original unit of work is neither committed nor in-doubt, discarding the cached operations.

8. A method according to any one of the preceding claims, wherein data restored to the primary copy of the repository within the recovery unit of work is made inaccessible by setting a flag for each data item restored to the data repository, the flag indicating that the data item is not accessible.

9. A method according to claim 8, wherein the flag indicates a transactional state of the data item and wherein a process for retrieving data items from the repository is adapted to identify one or more predefined transactional states as inaccessible.

10. A method according to claim 8 or claim 9, wherein the flag comprises a byte value of a distinctive primary key allocated to the data item when the data item is restored to the data repository, the byte value being selected from a range of values indicative of the transactional state of the data item.

11. A method according to any one of claims 8 to 10, wherein the step of setting a flag comprises:

setting a first flag for any data item for which the latest operation performed on the data item prior to the failure was a committed add operation which is to be restored to the data repository within the recovery unit of work; and

setting a second flag for any data item for which the latest operation performed on the data item prior to the failure was an in-doubt add or delete operation which is to be restored to the data repository within the recovery unit of work.

12. A method according to claim 11, wherein the first flag comprises a byte value of a data item key selected from a first range of byte values representing a first transactional state and the second flag comprises a byte value of a data item key selected from a second range of byte values representing a second transactional state.

13. A data communication system including:

data storage for storing a primary copy of a data repository;

secondary data storage for storing a secondary copy of data representing the data repository which secondary data is sufficient to recover the primary copy of the data repository and data held thereon;

a recovery component for controlling the operation of the data communication system to recover from a failure affecting the primary copy of the data repository, wherein the recovery component is operable to control the data communication system to perform the steps of:

recreating a primary copy of the data repository from the secondary copy; and

using a restore process to restore data items to the primary copy from the secondary copy within a recovery unit of work, wherein data items restored to the primary copy of the data repository within the recovery unit of work are made inaccessible to processes other than the restore process until commit of the recovery unit of work;

prior to commit of the recovery unit of work, configuring the primary copy of the data repository to enable addition of data items to the data repository independent of said restore step and to enable processes other than the restore process to retrieve said independently added data items; and

in response to successful completion of the restore step, committing the recovery unit of work including releasing said inaccessibility of the restored data.

14. A data communication system according to claim 13, comprising a data communication system for transferring messages between a sender and a receiver, wherein messages are held in the data repository following a message send operation by the sender and the messages are subsequently retrieved from the data repository for delivery to the receiver, and wherein a backup copy of the data repository is created and log records are written to record message send and message retrieval events since creation of the backup copy;

wherein the recovery component is adapted to control the data communication system to perform the following steps:

in response to a failure affecting the data repository, restoring messages to the data repository by reference to the backup copy of the data repository which backup copy was created prior to the failure;

prior to completion of the recovery processing, configuring the data repository to enable new messages to be added to the data repository and retrieved therefrom without awaiting completion of the recovery processing; and

reapplying updates to the data repository corresponding to message send and message retrieval operations performed prior to the failure, by reference to log records created prior to the failure;

wherein the steps of restoring messages to the data repository and reapplying updates to the data repository by reference to the backup copy and log records are performed within a recovery unit of work and the restored messages and reapplied updates are made inaccessible until all data repository updates corresponding to send and retrieve operations performed prior to the failure have been reapplied to the data repository.

15. A computer program product comprising program code recorded on a recording medium for controlling the operation of a data processing apparatus on which the program code executes to perform a method for recovering a data repository from a failure affecting a primary copy of the data repository, for use with a data processing apparatus having a secondary data storage and having a component for maintaining a secondary copy of data in the secondary data storage which secondary copy is sufficient to recover the primary copy of the data repository and data items held thereon, the method including the steps of:

in response to a failure affecting the primary copy of the data repository, recreating a primary copy of the data repository from the secondary copy; and

using a restore process to restore data items to the primary copy from the secondary copy within a recovery unit of work, wherein data items restored to the primary copy of the data repository within the recovery unit of work are made inaccessible to processes other than the restore process until commit of the recovery unit of work;

prior to commit of the recovery unit of work, configuring the primary copy of the data repository to enable addition of data items to the data repository independent of said restore step and to enable processes other than the restore process to retrieve said independently added data items; and

in response to successful completion of the restore step, committing the recovery unit of work including releasing said inaccessibility of the restored data.

16. A computer program for controlling the operation of a data processing apparatus on which the program executes to perform a method for recovering a data repository from a failure affecting a primary copy of the data repository, wherein the data processing apparatus has a secondary data storage area and wherein the computer program includes a component for maintaining a secondary copy of data in the secondary data storage area which secondary copy is sufficient to recover the primary copy of the data repository and data items held thereon, the method including the steps of:

in response to a failure affecting the primary copy of the data repository, recreating a primary copy of the data repository from the secondary copy; and

using a restore process to restore data items to the primary copy from the secondary copy within a recovery unit of work, wherein data items restored to the primary copy of the data repository within the recovery unit of work are made inaccessible to processes other than the restore process until commit of the recovery unit of work;

prior to commit of the recovery unit of work, configuring the primary copy of the data repository to enable addition of data items to the data repository independent of said restore step and to enable

processes other than the restore process to retrieve said independently added data items; and

in response to successful completion of the restore step, committing the recovery unit of work including releasing said inaccessibility of the restored data.